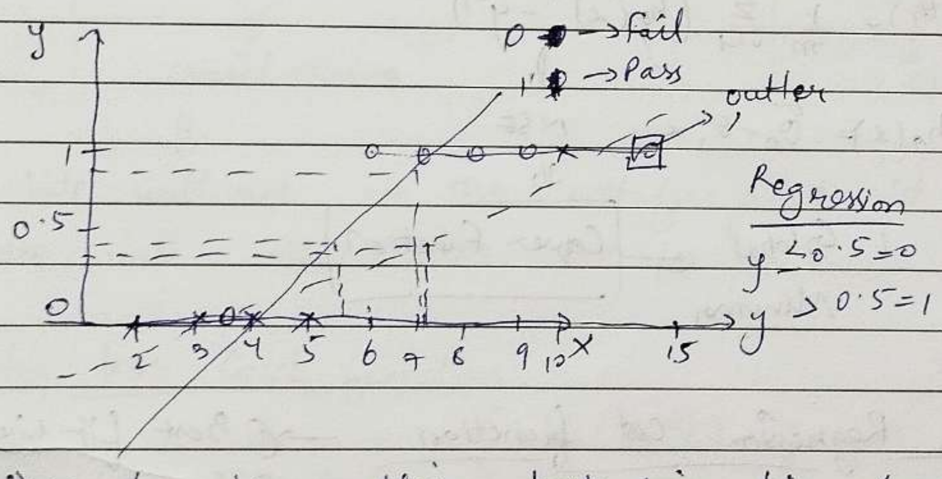
Logistic Regression (Classification Problem)

DATASET

Study hours	o/p (Pass/Fail)
2	FAIL
3	FAIL
4	FAIL
5	FAIL
6	PASS
7	PASS
8	PASS
9	FAIL

Q. Can we solve this using Linear Regression.



$0.5 \Rightarrow$ Threshold

Regression

$y < 0.5 = 0$

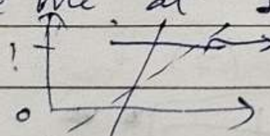
$y > 0.5 = 1$

→ Due to the outlier best fine line changes.

→ When we plot (7) value it is less than 0.5. ∴ person should value

→ If we plot for (17) it would be greater than 1.

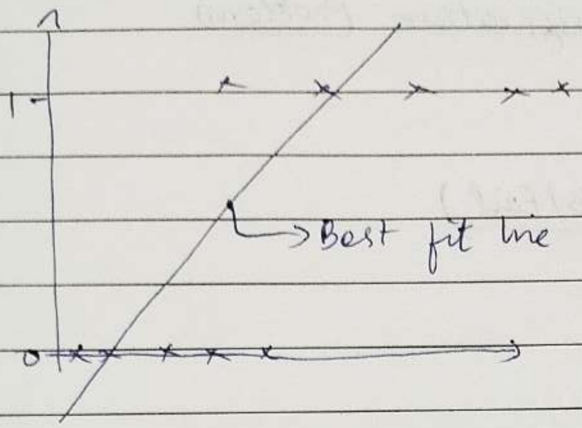
→ We have to squash the line at 1 and 0 & it is done using logistic Regression!



→ line should stop here

→ This squashing is done using a function Sigmoid Activation Function.

Sigmoid Activation



Best fit line $= \boxed{h_\theta(x) = \theta_0 + \theta_1 x}$

$\Downarrow$

Sigmoid Activation $\Rightarrow$ o/p = 0 to 1

1. $z = h_\theta(x) = \theta_0 + \theta_1 x$
2. sigmoid fn $= \boxed{\dfrac{1}{1+e^{-z}}} \Rightarrow$ 0 to 1

Linear Regression Cost function

$$J(\theta_0, \theta_1) = \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x)^{(i)} - y^{(i)}\right)^2$$

$$h_\theta(x) = \theta_0 + \theta_1 x$$

$\Downarrow$ MSE $\Downarrow$

1 Global Minima $\Leftarrow$ $\boxed{\text{Convex Function}}$

Logistic Regression cost function $\rightarrow$ 1. Best fit line
2. Squashing $\rightarrow$ Sigmoid function

$$J(\theta_0, \theta_1) = \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x)^{(i)}\ y^{(i)}\right)^2$$

$$h_\theta(x) = \sigma(\theta_0 + \theta_1 x)$$

$\Downarrow$

Sigmoid Activation

$\boxed{z = \theta_0 + \theta_1 x}$

$$= \sigma(z)$$

$$= \boxed{\frac{1}{1+e^{-z}}}$$

$$\boxed{h_\theta(x) = \frac{1}{1+e^{-(\theta_0 + \theta_1 x)}}}$$

$$\boxed{h_\theta(x) = \frac{1}{1+e^{-(\theta_0+\theta_1 x)}}}$$

$\Downarrow$

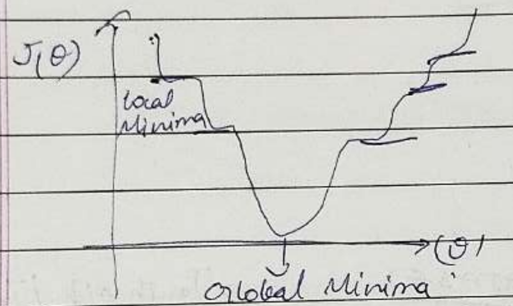
O/p $\rightarrow$ 0 to 1

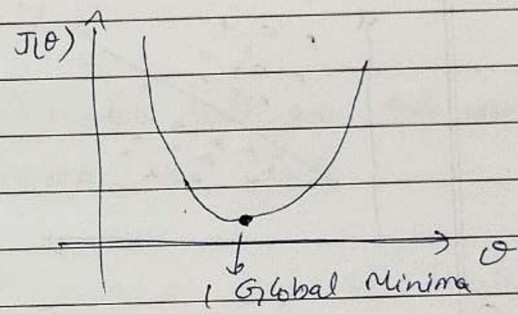
Threshold $\leq 0.5 \Rightarrow 0 \Rightarrow$ Fail

$\geq 0.5 \Rightarrow 1$ - Pass

This creates a non-convex function ~~Non~~ convex fn.

Non-Convex ~~Eqn~~ fn.





$\Downarrow$

We will not use this cost fn. since it creates non-convex fn.
We will use another cost fn.

log loss Cost function

$$Cost\left(h_\theta(x)^i, y^{(i)}\right) = \begin{cases} -\log(h_\theta(x)) & \text{if } y=1 \\ -\log(1-h_\theta(x)) & \text{if } y=0 \end{cases}$$

y → Truth Point

$$\boxed{Cost\left(h_\theta(x)^{(i)}, y^{(i)}\right) = -y\log\left(h_\theta(x)\right) - (1-y)\log\left(1-h_\theta(x)\right)}$$

$\Downarrow$

This will give convex fn.

Through this we never get a local Minima.

Minimize cost fn. $J(\theta_0, \theta_1)$ by changing θ_0, θ_1

Convergence Algorithm,

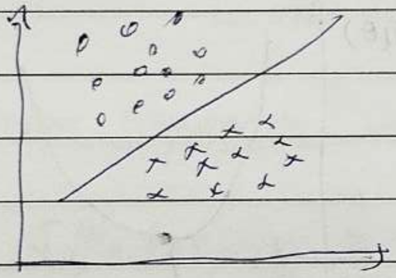
Repeat Until Convergence,

$\{$ $j=0$ and 1

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$$

$\}$

PERFORMANCE METRICS



DATASET

$y \to$ Truth o/p / Truth point

$\hat{y} \to$ Prediction

1. Confusion Matrix
2. Accuracy
3. Precision
4. Recall
5. F-Beta Score

f_1	f_2	O/P y	$\hat{y}$
–	–	0	1
–	–	1	1
–	–	0	0
–	–	1	1
–	–	1	1
–	–	0	1
–	–	1	0

CONFUSION MATRIX

	1	0
1	3	2
0	1	1

Columns: y Actual value; Rows: $\hat{y}$ Predicted value

$\Rightarrow$ 2×2 Matrix

	1	0
1	TP	FP
0	FN	TN

TP → True Positive
TN → True Negative
FP → False Positive
FN → False Negative

$$Acc = \frac{TP + TN}{TP + FP + FN + TN} = \frac{3+1}{3+2+1+1} = \frac{4}{7} = 0.57$$

* DATASET → Binary Classification.
↳ 1000 Datapoints :- → 900 (says 1), → 100 (says 0) } — Imbalanced Dataset

Dumb Model → (O/p as 1) — This model will give ~~o/p as 1~~ accuracy as 90%

⇓

This is the reason we cannot rely on confusion matrix.

* Precision :- $\frac{TP}{TP+FP}$

	1	0
1	TP	FP
0	FN	TN

Actual
Predicted

FP↓↓↓

Out of all the actual values, how many are correctly predicted.

Problem Statement
Mail → Spam / Ham

→ We need to reduce FP because if a Mail is ham & if it predicts as Spam then it is problem.

* Recall :- $\frac{TP}{TP+FN}$ ⇒ Out of all predicted values how many are correctly predicted.

→ Tomorrow the stock market is going to crash.

$\Downarrow$

→ Consumer → FN↓↓

→ Companies → FP↓↓

In this Scenario both FN & FP can be important.

$\Downarrow$

In this Scenario we use F-Beta Score :-

$$\boxed{(1+\beta^2) \times \frac{\text{Precision} \times \text{Recall}}{(\beta^2 \times \text{Precision} + \text{Recall})}}$$

(i) If FP & FN are both imp. then $\beta = 1$,

$$F_1 \text{ score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

(ii) If FP is more important than FN

$\beta = 0.5$

$$f_{0.5} \text{ score} = (1+0.25) \times \frac{\text{Precision} \times \text{Recall}}{(0.25 \times P + R)}$$

iii) If FN is more important than FP

FN >> FP

$$F_2 \text{ score} = \frac{(1+4)\, P \times R}{(4 \times P + R)}$$

Assignment

i) 90% Accuracy if data is not imbalanced
ii) If data is imbalanced without handling:
Precision, recall, F1-score

- If data is imbalanced handle imbalance data & create a model.